Spatial Inclusion and Analogy for Set Membership: A Case Study of Analogy at Work

Spatial inclusion and set membership: a case study of analogy at work

Keith Stenning Jon Oberlander

Human Communication Research Centre
University of Edinburgh*

Abstract

The analogy between spatial inclusion and set membership offers the opportunity to go beyond the study of how we establish analogical mappings to the investigation of the impact of those mappings on mental process. This analogy is the basis of longstanding didactic techniques for teaching elementary logic (Euler's Circles and Venn Diagrams) and therefore bears on psychological theories of human deductive reasoning (e.g. Erickson 1974).

An adequate cognitive theory of analogy must explain how mental processing is aided by analogical mappings. In the case of spatial inclusion as an analogy for set membership this comes down to requiring a theory of how graphical representations aid processing. Here we sketch a general theory in terms of the specificity of graphics which we have developed further elsewhere (see Stenning & Oberlander 1991a, 1991b). This general approach to graphical representations rests on the observation that graphical representation systems limit abstraction and thereby purchase computational tractability.

In applying this general theory to the case of graphical methods of syllogism solution we sketch a method of using Euler's Circles and show how interpretative conventions for graphical objects allow the necessary abstractions. In particular we show that even choosing the least expressive system of diagrams limited to circles, still gives the requisite expressive power.

The Euler's Circle method sketched is then used as a guide to connectionist implementation of corresponding internal representations. Graphical representations share some general computational properties with connectionist systems. In the domain of the syllogism, the central issue is the implementation of the binding of attributes into type-descriptions. Various proposals for connectionist variable binding in the literature are assessed as partial implementations. Some proposals are made for implementation of novel graphical features of the Euler's Circle algorithm.

This case study of an analogy suggests the hypothesis that one important role that analogy may play is providing a representation which is more limited in expressive power and therefore more computationally tractable than the literal representations analogies supplant.

*The support of the Economic and Social Research Council UK (ESRC) is gratefully acknowledged. The work was part of the research program of the ESRC funded Human Communication Research Centre (HCRC). We also thank Peter Yule for insightful suggestions of improvements to the graphical algorithm.

1 Introduction

An analogy consists of an object and something to which that object is likened. This comparison is made in order to understand or reason about the object, and the purpose of analogy is to facilitate these processes. Adapting a terminology of I. A. Richards (1936), itself based on an analogy, I will call the object of the analogy the *topic* and the thing it is likened to the *vehicle*. The vehicle of an analogy (at least a good analogy) enables or facilitates some reasoning about the topic. A cognitive theory of analogy must explain this facilitation as well as what mapping is established and how.

This paper is a case study of a particular analogy between spatial inclusion and set membership. While this analogy may have features which others lack, and may lack features which others have, as an example it has the virtue that it is possible to be rather precise about the properties of both vehicle and topic. Furthermore, because we can specify the reasoning which the analogy is supposed to facilitate, we can hope to be precise about the facilitation in reasoning that results from the employment of the analogy. Even if some of the empirical observations have not yet been made, it is possible to specify what would be required to test this theory of the efficacy of an analogy in achieving the facilitation of, say, syllogistic reasoning. Thus we intend a claim that this *is* a good analogy. The very staying power of graphical teaching methods in elementary logic is strong anecdotal evidence. We intend our theory to explain why these teaching methods work.

We did not initially set out with analogy as our object of study. The original focus was an account of differences between graphical and linguistic representations. But one result of this starting point is the realisation that analogies are representations. Vehicles represent topics, and operations defined on vehicles can then be interpreted as operations on their topics. Just as analogies consist of partial mappings—not every property of the vehicle has an interpretation in the domain of the topic, so representations have properties in virtue of which they represent, and other properties which are irrelevant to their interpretation as representations. Related views of analogy as representation have been put forward by Holland et. al. (1986), and by Halford, Wilson, Guo, Wiles & Stewart (this volume).

Shifting perspective on the problem in this way suggests the extension of a claim developed elsewhere as an explanation of why graphical representations aid reasoning. Our hypothesis is that graphical representations aid processing because they restrict abstraction (see Stenning & Oberlander 1991a and Stenning & Oberlander 1991b for more extensive development of this theory). Extended to a general hypothesis about analogy, this would suggest that analogies work because vehicles are less expressive (in a well defined technical sense of expressive) than their topics, and therefore easier to process.

A possible complaint against this approach to analogy is that our example is anaemic. If the reader's interest in analogy is the role it plays in reasoning about ill-formulated domains, or in learning about conceptually rich and novel domains, then this is not the obvious example to choose. Users of this analogy typically know all that they need to know about set membership before they encounter this analogy. It would be difficult indeed to decide whether vehicle or topic is epistemologically more basic, both being about as basic as concepts can get. Some might argue that ontogenetically the one is derived from the other. Piaget (1956) has argued thus, and the localist hypothesis (see e.g. Lyons (1968) for a review) which

claims that abstract concepts such as set membership develop from spatial relations such as containment has some linguistic support. This author is sceptical about this hypothesis as an *epistemological* derivation for Kantian reasons: it is hard to see how individuals, or space or time or any other fundamental concept can be independent of the basic concepts of set theory. As a *developmental* derivation, the hypothesis may have much to recommend it. Indeed the developmental version of the localist hypothesis amounts to the theory that our example analogy operates during ontogenesis. Be that as it may, the average introductory logic student may be assumed to be beyond Paiget's stages, and the efficacy of the analogy for *calculating* syllogistic conclusions remains after conceptual developments are complete.

Another source of complaint about our example is the temptation to etymologise ana*logy* as meaning 'without logic' (see, for example, Johnson-Laird (1983) where analogical reasoning is contrasted with logical reasoning). Our analogy is actually a palpable aid to logical reasoning of the most mundane variety—quite fitting for a word whose actual etymological derivation is from a mathematical term for equality of ratios.

There are two responses to these complaints about our choice of example. We do not mean to impugn the importance of analogy as a method of reasoning in ill-defined conceptually rich domains. Ours is a modest proposal not intended to do more than throw some light on one aspect of a large topic, an aspect that may be usefully studied in a domain where vehicle, topic and reasoning task are all well defined.

The second response is that even this anaemic analogy appears to be capable of playing a role in bringing about rather fundamental changes in people's grasp of some of the most abstract of concepts. Part of the problem novice logic students face is understanding the difference between validity and truth. We will argue that our example analogy plays an important role in making some of the most abstract aspects of model theory transparent to language users who have not previously much engaged the exercise of pure deductive inference. If such an anaemic analogy can have this significant conceptual impact, it may prove more full-blooded than we at first supposed. It may also get us to reconsider our etymologies, and the relation between logic and thinking.

The plan of the paper is as follows. In the next section we briefly outline the theory of graphical representations that lies behind this approach to investigating their role in mental processes. In Section 3 we sketch a method for using Euler's Circles to draw inferences from pairs of syllogistic premisses. In Section 4 we then examine in more detail the correspondences (and non-correspondences) between our graphical vehicle and its logical topic, measuring the general theory against the details of the example, illustrating how the graphical representations limit the expressive power of language. In Section 5 we ilustrate how graphics still permit just the abstractions that are required for Euler's Circles to constitute a theorem prover for this small fragment of elementary logic. In Section 6 we consider the relation between our graphical system and mental implementations of syllogistic reasoning. Finally, in Section 7, we return to draw some conclusions about analogy from this comparative study of linguistic and graphical representations.

2 Comparing graphical representations with language

One goal of a cognitive theory of graphical representation is to explain why graphical representations have different cognitive characteristics than linguistic representations. At a pre-theoretical level, the phenomena to be explained are similarities and differences between performance with *external* graphical and *external* linguistic representations. We want a theory which will explain why people perform as they do with the same tasks posed in linguistic, in graphical and also in mixed linguistic and graphical terms. Our intuitions are that many inferential tasks are just obviously more easily posed and solved in graphical terms though some may actually be harder. A satisfactory theory will explain both these types of case. We acknowledge that there are many difficulties in equating problems expressed in different media. This is an important part of the reason for wanting a theory of graphical representations couched in logical/computational terms. Nevertheless, for an excellent discussion motivating the intuition we recommend Barwise and Etchemendy (1990)

A cognitive theory of performance with external graphical and linguistic representations will not be able to avoid giving a theoretical account of the *internal* representations which support performance. When we come to discuss implementation, we will point out some affinities between graphical representations and connectionist mechanisms. We see as one of the main benefits of formulating a general theory of the logical and computational properties of graphics, the clarification that results when considering questions of internal implementation of the structures and processes that reason over the external representations.

We will not do more than briefly describe our general theory here insofar as is necessary to understand the specific example analogy at hand. The central tenet of the theory is that graphical representations limit the expression of abstractions and thereby improve processability. The intuition that graphics limit abstraction is at least as old as Bishop Berkeley's arguments against Locke's picture-theory of word meaning (Berkeley 1710). A picture of a triangle has determinate angles and ratios of sides, whereas the word 'triangle' fixes none of these. The word abstracts where the picture cannot. This insight needs to be related to the equally often expressed intuition that actual systems of graphical reasoning do manage to express abstractions graphically. The geometry diagrams that concerned Berkeley *do* express abstractions when one analyses their deployment in the systems of inference in which they are embedded. A proof must make no recourse to the 'accidental' properties of its example diagrams. By the same store, examples are chosen to be the most general cases—an irregular triangle rather than an equilateral one if the proof is not to depend on equal sides.

These twin intuitions appear to give with one hand and take away with the other. Graphics limit abstraction in some ways, but allow it in others. This impression appears to us to be the main reason why these twin intuitions have never been built into a general theory of graphical representation. On close examination, this impression of a 'conservation of abstraction' can be seen for the conjuror's illusion that it is. What the concreteness of 'primitively' interpreted graphics takes away with one hand is not the same as what the employment of abstractive conventions of interpretation give with the other. Watch very carefully and you will see the rabbit left behind in the hat.

What is required to turn these twin intuitions into a theory is first a precise account of what limitations on the expression of abstraction primitively interpreted graphics impose, and

second a precise account of what abstractions can and which cannot be captured by abstractive conventions of interpretation working over these graphics. The difference is what makes the theory contentful. If all abstractions that were denied by graphics could be reinstated with equal facility by conventions of interpretation, then we would be left with no theory of the cognitive differences between these two types of representational system.

A logical characterisation of the concreteness of graphical representations can be given by defining languages with limited powers of abstraction. One of the most powerful limitations on the logical language of graphical representations is that they contain no quantifiers and the identity relations between all constants are fixed by axioms of identity. This reflects the fact that it is not possible to picture two elements without determining whether they are the same element or not, and is one of the most powerful aids to making inferences from graphical representations. The remainder of graphical specificity consists in the fact that the spatial vocabulary comes in sets of items which mutually determine each other. It is not possible to fix a relation such as 'x is right of y' in a picture without also fixing whether or not x is left of y. This aspect of specificity is far harder to exhaustively capture logically, but nevertheless corresponds to the extreme overdetermination of graphical representations (see Stenning & Oberlander (1991b) for a fuller account where these ideas are related to those of Levesque (1988) and others). The reader will have realised from this description that specificity is intended as a property of representation systems, not of single representations themselves. What distinguishes pictures from language is that language can determine all or none of these features of a depicted scene. Pictures force us to determine them whether we care to or not.

What abstractions can be made by conventions of interpretation for these concrete graphical objects? How can one diagram be made to stand for a disjunction of possibilities? Here it seems that the line that can be drawn is not so hard and fast. We can state ever more complex interpretative conventions for some graphical system of representations, which allow ever more abstractions to be expressed. However, we have a clear sense that more and more work is being done by the language and less and less by the pictures that the language helps us interpret. We are confident that careful empirical investigation of the efficacy of such complexly interpreted graphics would show that the pictures are not worth the thousands of words they require for their interpretation.

If we cannot yet give a full account of what abstractions can easily be achieved by interpretation conventions, we can indicate some clear guidelines. Monadic properties of pictured elements are easily abstracted over. This is what happens when we define 'symbols' in a diagramming system or icons for an interface. We understand that squares stand for zebras, or items costing more than $1000, or whatever. These symbols function very like words, and their crucial property is that their semantics is not internally graphically componential. At this level, it is quite easy to have a triangular icon function very like the abstraction expressed by the word 'triangle' because the graphical properties of the icon are unanalysable—the angles are not there to be interpreted as part of the diagram save through their status as part of the icon.

A further step towards abstraction in the interpretation of symbolic icons is through interpretative conventions which treat icons as meta-objects. So, for example, a common convention is to designate a particular shaped icon to operate as a variable ranging over a domain of shapes rather than as representing its own shape. A cylinder, say, then means an object of

any shape (excluding a cylinder, which shape becomes unrepresentable). Of course it would be possible to allow this cylindrical icon to represent cylinders as well, but this seems a less intuitive convention for understandable reasons. A cylinder icon still cannot be interpreted as standing for a cylinder since it might stand for any other shape. In logical terms, it is useful to be able to distinguish between variables and constants.

Again such conventions operate only for monadic properties. The difference between this meta-object icon (cylinder-for-any-non-cylinder-shape) and an exemplar icon (specific-triangle-for-any-triangle) appears to be more than one of degree. This can be seen by comparing this meta-object icon convention with one in which the cylinder stands for cylinders *as well as any other shape*. This later type of convention is merely a more general case of abstraction-from-exemplar like the specific-triangle-for-any-triangle convention. But it intuitively has rather different psychological properties from the meta-object convention—namely it is much more confusing. It would be interesting to know whether these intuitions can be substantiated.

Matters change dramatically when we consider polyadic properties (either by analysing a symbol into its graphical parts, or by considering relations between symbols). Suppose we want to enter a symbol for a pawn into a picture of a chess board but we want to abstract over a range of positions. If we have one pawn on a chess board, we might try defining a convention which said that its position was only to be thought of as determinate with regard to the half of the board it is on. This is fine until we enter another piece. The two pieces then automatically have a whole range of their mutual spatial relations fixed in the picture. But if we are to maintain our abstractive convention, only some of these relations will be interpretable. So, if they are in the same half of the board, this fact can represent itself. But within that half, the fact that X is nearer the edge than Y cannot be interpreted to mean that this is true of the actual pieces. Matters degenerate still further if some pieces' positions are deemed to represent their positions, and others not. One might as well have a linguistic data base of the known facts about the pieces' positions as a misleading picture and a very large list of interpretative conventions telling us which graphical relations can be interpreted and which not. In fact the data base would look remarkably like the list of conventions.

These remarks fall far short of a logical characterisation of the possibilities for abstractive conventions but they suffice to show that nothing like the resources of a polyadic quantified language can be attained by interpretative conventions of reasonable complexity for graphical representations. They also strongly suggest that the interesting cognitive properties of graphical systems will emerge from consideration of their portrayal of relations rather than of the behaviour of iconic representation of word-like symbols. Our analogy between set membership and spatial containment falls in exactly this domain.

The role of graphics in reasoning is a topic with a long history in AI. Gelernter (1963) reports work from the late 1950s which used geometry as a domain in which to show that giving a diagram to a heuristic theorem prover provided opportunities to drastically prune the search space for proofs. Simply rejecting subgoals which were false in the diagram was enough to bring previously intractable proofs within the ambit of their theorem prover. These techniques rely on statistical properties of geometry diagrams. Any particular diagram is sampled from an extremely large space of possible diagrams. The chances, for example, that two lines in an arbitrarily constructed example diagram should be equal unless they are equal in all

possible diagrams is quite small. Gelernter specifically notes that this technique has problems with proving inequalities where this reasoning cannot help. The chances that an arbitrarily constructed example diagram should have AB≠CD are high, whether or not this is true of all examples i.e. a theorem.

Within the framework developed here, Gelernter's work is a good example of how the employment of graphics in reasoning aids processing by curtailing abstraction and how conventions of interpretation can achieve the expression of some abstractions but not all.

Funt (1977, 1980) developed this line of research further by incorporating simulations of the perception of diagrams into a system for reasoning about their mechanical properties. A retina of communicating elements with a simulated parallel inference mechanism made rapid detection of spatial properties such as the centers and symmetries of planar solids. A high level reasoner took these judgements as input and made inferences about the trajectories of fall of shapes in unstable stacks of blocks. This is a problem which would be hard for a general purpose reasoner working on data expressed as propositions in a fully abstractive language. It is the spatial nature of the retina which enables rapid extraction of spatial properties from the input. This work underlines what Stenning and Oberlander take as given. The efficient content addressable search of graphics for data for reasoning is possible because our visual system can perform these low-level inferences using parallel mechanisms. But Funt takes for granted what Stenning and Oberlander emphasise—that it is the lack of abstractive power of diagrams (and scenes) which means that our visual system could evolve. If our input were in full first order predicate calculus, there would be no such parallel processing system. Both Funt and Gelernter can be seen as much more detailed workings out of the current general approach within the particular domain of geometrical reasoning.

Other work has had goals closer to those of Stenning and Oberlander's general account comparing reasoning with graphical and with linguistic representations. Lindsay (1988) contrasts 'deductive' reasoning with the sort of non-discursive reasoning that graphics makes available. Once input has been organised as an image, results can be just read off. Biological systems, which give a high priority to rapid response but may be content with slow 'off-line' input, favour this design choice. Learning a skill may be slow, but its excercise generally requires fluency. The motivation of Lindsay's argument is similar: the detail differs in that he relies on an architectural distinction to define his rather non-standard sense of deduction, and he lays less stress on the curtailment of abstraction that image construction enforces.

Finally, Johnson-Laird's (e.g. 1983) work on mental models is an approach to spatial reasoning and therefore related to our current concerns. Mental models have some of the properties of graphical representations—they are two dimensional and are agglomerative in that they enforce connectedness between represented elements. They amount to a strategy of theorem proving which proceeds by building models (of the logical variety) which initially may force over-specification. However, the exposition Johnson-Laird chooses stresses the idea that mental models are 'analogical' and semantic, and disguises their connection to prior graphical systems such as Euler's Circles and indeed their relation to logical methods such as 'tableau' proof systems. Neither does Johnson-Laird give any account of what limitations on expressive power mental models have and therefore no account of how this aids inferential tractability. We will see below that mental models are much more closely related to Euler's Circles than Johnson-Laird allows, but that Euler's Circles afford efficiencies of memory and reasoning by

exploitation of their graphical nature.

3 Euler's Circles as an extended analogy for syllogistic logic

The analogy between the spatial containment of points and the membership of elements in sets is actually, like most interesting example analogies, only the iceberg tip of an extended analogy. The intersection of sets is analogous to the overlapping of regions. The set—proper subset relation is analogous to the containment of one region by another. Separate regions are analogous to disjoint sets. Relative movement of pairs of regions corresponds to ranges of set relations. Constraints on movement can be interpreted as logical relations between sets.

We first describe this extended analogy as it functions in one method of using Euler's Circles to implement syllogistic logic. We beg the reader's patience while we specify the details of this particular development of our analogy. The detail is necessary for an understanding of how the analogy *works*. A more detailed treatment of this graphical method and its relations to other methods of syllogistic reasoning is given in Stenning & Oberlander (1991b).

We prefer to develop this highly procedural account of the use of Euler's Circles rather than analyse the version of the analogy embedded in the use of Venn Diagrams. The latter are probably more widely used in logic text books and they have some virtues for logical analysis. However, they are much less thoroughly graphical in that they exploit extensive annotation systems added to a single graphical arrangement of circles. It is our intuition that Euler's Circles have much more interesting psychological properties which stem from their greater graphical involvement, and its impact on human working memory, though we know of no empirical comparison of the properties of the two systems as teaching aids. Euler's Circles are a more interesting development of our analogy.

Some psychological theories of reasoning have hypothesised that people use a 'mental' version of Euler's Circles to solve syllogisms (e.g., Erickson 1974) but they have chosen to develop their theories with an idiosyncratic interpretation of the diagrams. This is probably because there are few if any explicit accounts of the use of Euler's Circles available in the literature. We sketch such a method here and refer the interested reader to Stenning & Oberlander (1991b) for further details.

A brief description of categorial syllogisms is in order. They are deductive arguments consisting of two premisses and a conclusion, but for our purposes are conveniently identified merely as premiss pairs. Premisses have one of four quantifier types relating three terms (e.g. All A are B; Some B are A; No C are B; Some C are not A). Pairs of premisses must share a 'middle' term (we choose 'B') and hence there are four *figures* (AB, BC; AB, CB; BA, BC; and BA, CB). The last can be transformed into the first by reordering the premisses (which does not affect the logic, but may affect the psychology). Conclusions contain the same four quantifiers but always relate A and C. Four quantifiers combined with two premisses in four figures yield 64 pairs of premisses. Of these 27 have valid conclusions.

The three predicates of a syllogism define eight types of individual: ABC, AB¬C, A¬BC, A¬B¬C, ¬ABC, ¬AB¬C, ¬A¬BC, and ¬A¬B¬C. Syllogisms are interpreted on domains consisting of selections from this set. Since syllogistic logic is a fragment of monadic logic with no identity relation, it is logically immaterial how many of a type of individual are in

a domain because the logic has no relational resources for distinguishing between individuals which have the same properties. Domains are also constrained to contain some As, some Bs and some Cs (the No-Empty-Sets Axiom). Furthermore, the inclusion or exclusion of the totally negative type ¬A¬B¬C is syllogistically immaterial—this type corresponds to the background of graphical representations. There are therefore 2^7 models minus those containing empty-sets relevant to the interpretation of the syllogism.

The strategy adopted by the theorem prover we describe is to represent only types of individual consistent with the syllogism's premisses. The range of types constructed represents both premisses in a single agglomerated representation. Having represented a range of consistent types, the method identifies any types of individual whose existence is established by the premisses. If there is no such type, there is no conclusion. If there is, then the process of formulating a conclusion revolves around this single type of individual. Existential conclusions may be drawn immediately; universal conclusions require checking of constraints.

The method we describe is a rational method which will yield correct conclusions in all cases, and which bears a particularly transparent relation to the underlying model theory of the syllogism. However, it can readily be related to the variety of error behaviour which subjects frequently exhibit.

We believe that psychologists' failure to understand how Euler's Circles are used is a result of adopting a primitive interpretation of the graphical representations. That they should have done this is a point of some interest to our theory of graphics and their abstractive interpretation to which we return below.

The Erickson interpretation of Euler's Circles is that each sub-region of a diagram stands for a type of individual which exists. This interpretation means that the mapping from diagrams of two sets to syllogistic premisses is many-to-one (see Figure 1). A more normal interpretation of these diagrams is that each region of a diagram stands for a type of individual which is *consistent* with the premisses (see Figure 2). Coupled with the strategy of representing all consistent types of individual this convention greatly simplifies the circles' use. The distinction between the two conventions can be brought out by adding to the latter an ancillary convention which shades any region representing types of individuals entailed by the premisses. This combined convention for interpreting Euler's Circle diagrams restores the one-to-one mapping of Euler's Circle diagrams to premisses. Probably the most perspicuous way of thinking of the shading convention is that shading distinguishes between derived (shaded) and assumed (unshaded) types. Figure 1 and Figure 2 compare the two different conventions.

The most complete theory of human syllogistic performance is due to Johnson-Laird (e.g. 1983) and is couched in 'Mental Model' Theory. Johnson-Laird makes much of the contrast between Euler's Circles and Mental Models, and of the inadequacies of Euler's Circles. Stenning (1991) and Stenning & Oberlander (1991b) have shown that this argument rests on Erickson's misinterpretation of Euler's Circles (which Johnson-Laird adopts) and when a faithful account of Euler's Circles is substituted mental models are notational variants of Euler's Circles. The latter are however, naturally constrained as a notation by their geometry, as we discuss below. In these matters, constraint is a virtue. Euler's Circles are directly based on an analogy, in a way that Mental Models obscure. Johnson-Laird cannot *explain* why mental model notation (which is just another proof-theory) is psychologically preferable to any other logical approach.

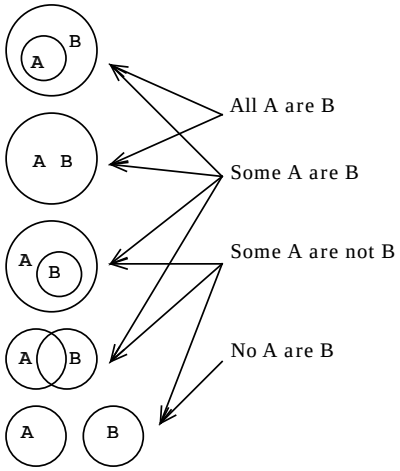

Figure 1: The five Gergonne relations between two circles mapped onto the four syllogistic premisses. Arrows point from a premiss to each of the diagrams in which the premiss is true.

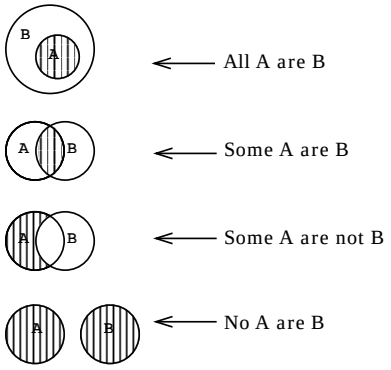

Figure 2: Characteristic Euler's Circle representations of syllogistic premisses. Regions represent types of individual consistent with premisses. Shaded regions represent types of individual established by the premisses.

With our interpretation of the regions, each premiss is now represented by a single shaded diagram, it remains to specify how the diagrams representing two premisses are 'registered' to yield a composite diagram, and how conclusions are drawn from this agglomerated representation. Figure 3 shows the eight registration diagrams which are the basis of all conventionally expressible syllogistic conclusions. In each case, the B circles of the two premisses are exactly superimposed. Positive and negative syllogisms are separated to emphasise that the nature of the logical constraints are different in the two cases (negative syllogisms are defined as syllogisms with at least one negative premiss). If only the B circles were labelled in the registration diagrams each diagram represents a group of syllogisms which can be derived by the two different assignments of A and C to the other two circles. Abstracting over these labellings reduces the number of diagrams and focusses on the relevant graphical properties. However, to aid reading, one arbitrary A/C labelling is assigned in each diagram. Figure 4 shows the diagrams for a group of syllogisms which license conclusions of the form Some not As are not Cs. These conclusions are not conventionally expressible in the syllogism. They also violate one of the principles of the conventional formulation of the syllogism that no pair of negative premisses have any valid conclusion. Figure 5 shows the registration diagrams for all other syllogisms.

In all registrations, the A and C circles are arranged with as much overlap as is consistent with the premisses—in other words the maximum number of types consistent with the premisses are represented. The existence of valid conclusion is identified with the existence of constraints on the consistent arrangement of the A and C circles. For positive syllogisms, some region exists which prevents A and C from being totally *separated*, and this region cannot be consistently eliminated. For negative syllogisms, some region exists which prevents A and C from being *superimposed*, and this region cannot be consistently eliminated.

The reasoner's task is to identify these critical regions. We begin by stating a normative strategy which is transparently related to the model theory of the syllogism. We will then look briefly at some of the heuristics which subjects appear to use and which are evidenced by their errors.

Shaded regions in the characteristic diagrams of premisses may be circular, crescent shaped or oval. A syllogism only has a valid conclusion if one of these shaded regions persists in the combined diagram *not intersected by any arc during the registration process*. Either the shaded region must be wholly included within the third circle derived from the other premiss, or it must be wholly excluded from that circle. In model theoretic terms, if C is the third circle and C intersects the shaded region from the premiss, then it is subdivided into two regions and no conclusion can be drawn because either one or the other of these regions can consistently be eliminated.

Our spatial analogy can be enriched into a mechanical analogy: if a nail is driven through one of these persistent shaded regions in the finished registration diagram, then it will constrain the movement of the A and C circles. In a positive syllogism, the nail will prevent the A and C circles from being slid apart; in a negative syllogism the nail will prevent the A and C circles from being superimposed. If no shaded region persists un-intersected in the registration diagram, then there is no valid conclusion.

Thus abstract properties of the model theory of the syllogism are directly reflected in geometric properties of circles. If these truths of model theory were parochial to the syllogism,

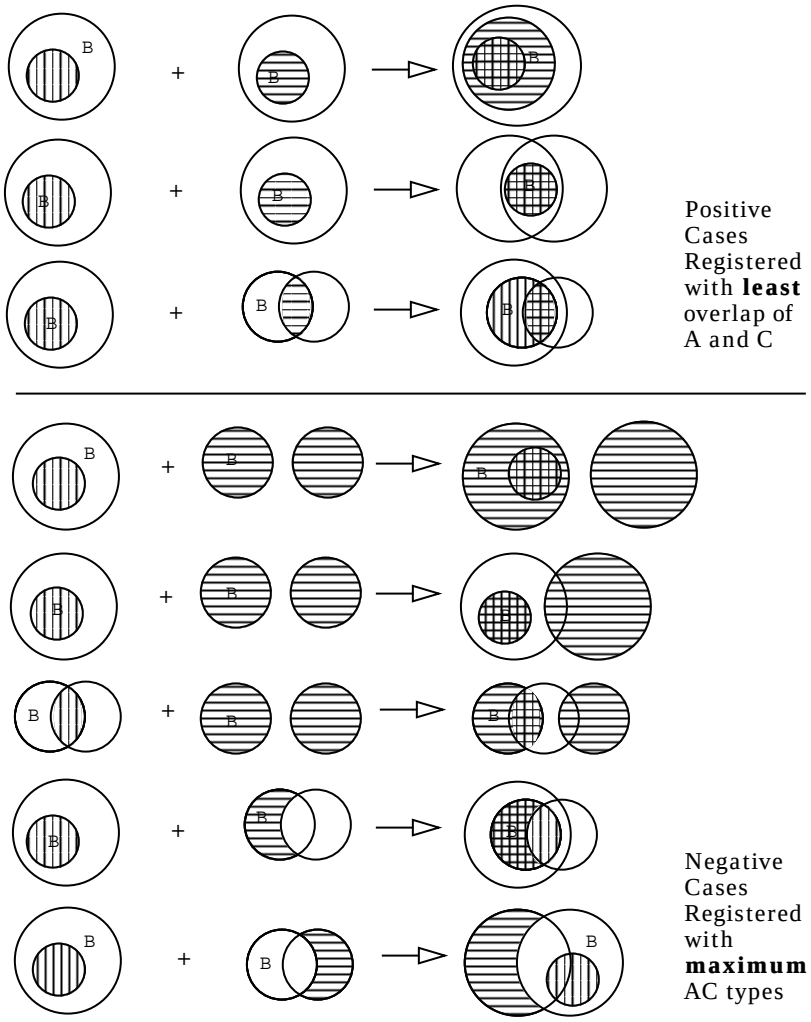

Figure 3: Registration diagrams with valid conclusions. (Key: premiss + premiss = registration diagram). Registration involves the superimposition of the 'B' circles of each premiss. The A and C labels are given for ease of reading—they can always be reversed throughout a registration sequence without logical effect.

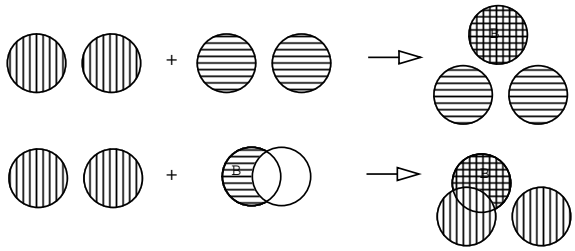

Figure 4: Registration diagrams with u-valid conclusions. (Key: premiss + premiss = registration diagram). These syllogisms warrant conclusions of the 'u'-form i.e. 'Some not A are not C'. N. B. The A and C circles can adopt any of the five Gergonne relations, but constraint lies in their mutual relation to B.

it would be hard to see how the spatial analogy could help—why should these stratgies of registration seem natural or indeed comprehensible. But these truths are not parochial, and students do appear to have some general grasp of them before approaching the syllogism. Their problem is to work out the consequences of their high-level grasp in the novel domain of arguments. We return to these points in Section 5.

The persistent shaded regions also play a central role in formulating conclusions—establishing the existence of a valid conclusion still leaves its linguistic formulation to achieve. Formulation of conclusions continues on from the identification of established individuals represented by the 'nailed' sub-regions of the diagrams. Existential conclusions can be drawn directly from the specification of these established individuals. Dropping the middle term predicate and existentially quantifying over the resulting conjunction of A or its negation and C or its negation yields the existential conclusion. So if the premiss establishes that there is an AB¬C, then conclude some A is not C.

Universal conclusions are more complex since they require that the nailed region of the registration diagram is circular and corresponds to either A or C. Intuitively, all members of A or C must be known to be of the established type in order to make a universal conclusion. In fact, only two registration diagrams (top positive and top negative in Figure 3) allow universal conclusions. Although the double shaded region in the second negative registration case down is circular, it represents B and therefore does not allow generalisation over all members of A or C, as the diagram shows. No other diagram in Figure 3 contains a double shaded circular area. In Figure 4, the only double shaded circular area corresponds to B, and so there is no call for a 'universal u-type' premiss.

A mechanical analogy corresponding to the nails for existential conclusions is that of a 'pastry cutter' for universal conclusions. Where there is a universal conclusion, such a device prevents the separation (in the case of positive syllogisms) or the intersection (in the case of negative syllogisms) of any part of the target region from the other circle. A universal generalisation

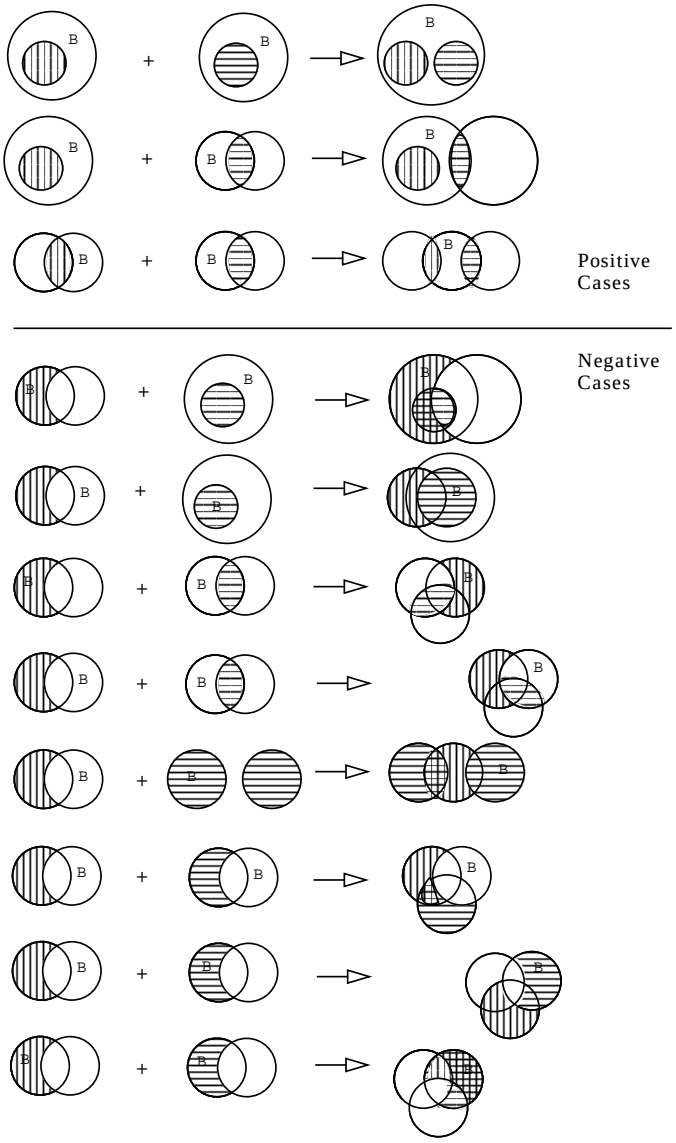

Figure 5: Registration diagrams with no valid conclusions. (Key: premiss + premiss = registration diagram)

about its contents is therefore valid. The pastry cutter wholly contains a circle within it whereas a nail merely guarantees or prevents intersection.

We have now described a system of reasoning which manipulates graphical, even quasi-mechanical, objects in such a way that a consistent interpretation of those objects in terms of set theory models syllogistic logic. In the next section we go on to consider what properties of the graphical objects and what properties of the logic permit this analogy, but first we make some comments on the relation between this normative algorithm and peoples' observed reasoning behaviour.

As the reader will probably have noticed, there is quite a close relation between the existence of areas of the registration diagrams which are doubly shaded and the existence of valid conclusions. The region which persists unintersected from premiss diagram into registration diagram tends to be double-shaded in the final result. In all but one of the eight registration cases which yield conventionally valid conclusions, there is an area of double shading once the two premiss-diagrams are registered, the exception being the bottom case in Figure 3. This syllogism is the hardest one with a valid conclusion in the data of several experiments on syllogisms (e.g. Johnson-Laird & Steedman 1978). The two u-valid registration diagrams also have doubly-shaded areas.

There are diagrams with doubly shaded areas which do not have valid conclusions (See Figure 5.). If subjects were simply using the existence of such areas to decide whether there is a valid conclusion these cases would invoke inferences, but many do not. However, there are easy clues to the absence of valid conclusions in many of the cases where there is double shading and we do not need to suppose that double-shading is used as the only basis on which to decide whether there is a conclusion. Firstly, there are two cases with double shading in which the registration diagram represents all seven possible types of individual (the bottom case and the case third from the bottom in Figure 5). If all types are possible, there is no valid conclusion. Secondly, these seven-type diagrams, and one other with double shading are also all cases in which two existential premisses combine and it seems likely that subjects have some explicitly formulated knowledge that there are no conclusions in these cases. It is interesting that this logical principle is robust, whereas the other of Aristotle's main principles (that no two negative premisses have a valid conclusion) is quite arbitrary, as we have seen. This leaves only positive cases and the top negative case in Figure 5. The three positive cases all exhibit double shading and all are observed to invoke quite high rates of invalid conclusion. The remaining negative case with double shading is the hardest no valid conclusion (NVC) syllogism of all—the presence of double shading appears to cue subjects to believe there is a valid conclusion.

As we shall see in Section 4, premisses with valid conclusions all entail the existence of maximal types of individual, maximal in the sense of being specified with regard to all three predicates. The shading convention, and our mechanical extension of the geometric analogy to 'nailing' is an expression of this logical property—the regions which are critical to reasoning are ones that represent individuals that must exist. Shaded regions of characteristic premiss diagrams represent types of individual specified on two properties. Doubly shaded regions of registration diagrams represent types of individuals specified by three predicates. This pattern of responses provides evidence that subjects are paying attention to these critical individual types in their reasoning in a way that the shading convention mirrors.

Further evidence is provided by performance in an extended syllogistic task in which subjects are asked to describe types of individuals which are entailed to exist by the premisses (Yule 1991). Although the u-valid cases are quite difficult, about half of subjects can correctly infer the existence of individuals which are B but are not A and not C. Given a task that allows subjects scope reveals their ability to reason in terms of maximal types.

To summarise this digression on the relation between our graphical algorithm and observations of human reasoning, there is a good correspondence between strategies revealed by the algorithm and observed performance. In particular, the presence/absence of double shading in registration diagrams, in conjunction with other factors, exerts a powerful influence in subject performance. The graphical algorithm suggests new observations which yield corroborating data. This is not the place for an extended treatment of the extensive empirical data on human syllogistic reasoning, but these preliminary observations are certainly suggestive.

4 The Inexpressiveness of Circles and the Limitations of Syllogistic Logic

Models of the syllogism are sets of types of individual drawn from the seven types: ABC, AB¬C, A¬BC, A¬B¬C, ¬ABC, ¬AB¬C, ¬A¬BC, and ¬A¬B¬C. The set {A¬BC, ¬A¬BC} is, for example, a model of the premisses 'Some A are not B. Some C are not B.' Having sketched one graphical method for reasoning about syllogisms we can ask what distinctions between models the system's representations can express, and what distinctions between models it is necessary to express in order to capture the logic of the syllogism. How do the geometrical properties of circles constrain distinctions between sets of types of individuals defined by their intersections? How is the expressive power that remains despite these constraints related to that necessary to capture the logic of the syllogism? Our hypothesis is that the benefit of our analogy is that it constrains the processing required for inference. We are therefore looking for gaps in expressive power of graphics which correspond to distinctions which are not necessary to capture the limited logic.

What models can Euler's Circles express? As we have seen above there are 2^7 models to consider. Some of these have empty sets (in which there are no As, or no Bs, or no Cs) and are thereby excluded from the standard interpretation of the syllogism. How do the graphical properties of circles constrain the expressive possibilities? Limiting oursleves to circles is the most restrictive graphical option. Other weaker constraints would be to allow convex regions, or closed curves in general, or even discontinuous curves. Each of these would allow more models to be expressed. However, in the spirit of our hypothesis that graphics aids reasoning by limiting expressive power we will opt to analyse what circles allow. We will see that circles are in fact quite sufficient to the logic of the syllogism.

Table 1 shows the number of models which have empty sets, the number without empty sets and having no Euler diagram, and the number without empty sets and having an Euler Diagram. These numbers are all under the restriction of the diagrams' vocabulary to circles of varying size. The models are classified by the number of elements which they contain. For instance, the model {A¬BC, ¬A¬BC} has two elements. About half of the models without empty sets are expressible. A rather smaller proportion of larger models are expressible than

No. of elements in model	Contains Empty Sets	No Euler Diagram	Has Euler Diagram	Totals
0	1	0	0	1
1	6	0	1	7
2	9	3	9	21
3	3	13	19	35
4	0	20	15	35
5	0	12	9	21
6	0	3	4	7
7	0	0	1	1
Total	19	51	58	128

Table 1: Numbers of models expressible by Euler's Circles classified by the number of their elements

of smaller models. Which models do Euler's Circles need to be able to express in order to capture the syllogism? Which distinctions between models are unnecessary? Syllogisms are highly constrained, not merely in their limitation to three predicates, but also in their quantificational/connective resources, and the limitation to a pair of premisses. As we have seen, of the seven logically relevant types of individual ($\neg A \neg B \neg C$ never makes a logical difference in the syllogism), one type cannot be the basis of any conclusion expressible by the four conventional quantifiers ($\neg AB \neg C$ requires a 'u' conclusion 'Some not A is not C'). Inspection shows that another type (A ¬ BC) cannot be established by any pair of syllogistic premisses (it requires a quantifier such as 'All not A are not B').

The best characterisation of the model discriminations necessary for capturing syllogistic logic can be given in terms of the relation between establishing types of individual defined by the A and C properties (or their negations) and the existence of valid conclusions. The syllogism has the very restrictive property that only pairs of premisses which establish the existence of such individuals have valid conclusions. Table 2 gives examples of syllogisms which establish each type of individual. For example, if all A are B, and all B are C, then there must exist individual(s) of the type ABC and this syllogism is one with valid conclusions. In contrast, if all A are B and all C are B, none of the seven types need exist (adequate models of the premisses can leave out each of the seven types) and there are no valid conclusions. It is this model-theoretic perspective that reveals the syllogisms diagrammed in Figure 4 that do establish the type ¬AB¬C but have no conventionally expressible conclusion (because the quantifier resources can't express the fact that there are some ¬As that are ¬Bs). Once an established type has been identified, the existence or non-existence of other types is logically irrelevant.[1] It is this property that licenses the focus of attention on a single type of individual and greatly reduces the need for distinguishing models, particularly large models. The inexpressible large models are inexpressible because it is impossible to express the exclusion of some types consistent with the premisses. This exclusion is always logically immaterial: the conclusions that do hinge on the existence or non-existence of these types are inexpressible in the syllogism. In monadic predicate calculus, every model can be uniquely specified

[1] This is a property of categorial syllogisms but *not* of logics in general. For example, from $\forall x(Ax \vee Bx), \forall x(Bx \vee Cx)$, *and* $\forall x(\neg Bx \vee \neg Bx)$, it follows that $\forall x(Ax \vee Cx)$, but none of the possible types of individual are established to exist.

Type	Example syllogism that establishes Type
ABC	All A are B. All B are C.
AB¬C	All A are B. No B are C.
A¬BC	None. (Require 'All ¬A are B').
A¬B¬C	No A are B. All C are B.
¬ABC	No A are B. All B are C.
¬AB¬C	No A are B. No B are C.
¬A¬BC	No A are B. All C are B.
¬A¬B¬C	None.

Table 2: Types of individual defined by three predicates with example syllogisms which establish their existence.

up to isomorphism. This result follows rather directly from the earlier mentioned one that numbers of types occurring in a model of monadic logic are logically irrelevant. For every type a conjunctive sentence can be constructed which is true if the type is in the model and otherwise false, so a conjunction of these conjunctions can be constructed which characterises uniquely each possible model (since the vocabulary is finite and the number of possible types is 2^p where p is the number of predicates in the vocabulary). The reader is referred to any introduction to meta-logic such as Hunter (1971) for further amplification of these properties of logics.

The graphical resources of Euler's Circles can be extended beyond the syllogism in some interesting ways. The type ¬A¬B¬C constitutes the background on Euler's Circles are drawn. In order for Euler's Circles to express propositions about the existence or non-existence of this type, Euler's Circles must be augmented by a device for distinguishing the domain set—a convenient one is to represent it by a square. The three predicates of a syllogism then may or may not exhaustively cover the square. Note that once the domain of reasoning is distinguished from the universal domain it becomes a live issue whether or not there are not As which are not B. The lack of the concept of domain is presumably what lead Aristotle to ignore U-type conclusions.

Earlier accounts of the limitations of graphical methods have noted the difficulty of extending Euler's Circles to more than three predicates (see Lewis Carroll's discussion of Venn's five predicate diagram in Dodgson 1896). They have not measured the expressiveness of Euler's Circles against the model theory of the syllogism, nor against its proof theory. What we have shown here is that if one takes the semantic space as the benchmark, Euler's Circles are highly constrained in what they can express. However, the syllogism is even more constrained and the constraints of the two types of representation are nested: Euler's Circles can express all the distinctions necessary to capture the syllogism.

In assessing our hypothesis about analogy, it is important to see that there are these three systems involved: the space of models, the graphical representations and the limited linguistic representations of the classical syllogism. We begin with a grasp of the fragment of naive set theory which structures the models, and an understanding of spatial containment, but initially we lack any further grasp of the syllogism beyond an understanding of the quantifiers and connectives involved. Spatial containment, and more particularly the spatial containments of overlapping circles on a plane, models only a part of the semantic possibilities. But it turns

out to be enough and more than enough to capture the very limited relations between sets which can be expressed in the novel logical fragment. The mapping is therefore helpful in reducing the space which we have to learn about.

5 How do Euler's Circles achieve abstraction?

We now return to our general approach to a cognitive theory of graphical representations and ask how this system of graphical reasoning achieves the abstractions required to capture syllogistic logic.

The pivotal shift from a primitively concrete interpretation of the diagrams to an abstract one is the shift from interpreting areas as corresponding to established types to interpreting areas as corresponding to consistent types. This is the first prerequisite to achieving a one-to-one mapping between diagrams and premisses and hence eliminating the disjunctions of diagrams necessary under the primitive interpretation. Note that this abstractive interpretative convention is global. It applies equally to all areas of a diagram, not just some of them. One can construct bizarre conventions which apply the abstractive convention to some areas and not to others (perhaps augmented with some extra colouring or annotation convention to distinguish the domain of application) but it seems intuitively clear that such conventions rapidly become just that—bizarre. The psychological complexity of their application outweighs any advantage that the use of graphics might confer.

Abstraction is available, but not just any abstraction. It is psychologically interesting that psychologists chose the interpretation of Euler's Circles which is primitive in our sense of the term—under Erickson's interpretation, one diagram stands for one actual model. Under the interpretation normally adopted in teaching logic, one diagram stands for many actual models but only one maximal model. Thus graphics are adapted to the logic by complicating the ontology. Such shifts are quite common in graphical systems, but we still claim that it is theoretically illuminating to distinguish between primitive interpretations and these more complex abstractive ones, and psychologists opted for primitive interpretation.

The addition of the shading convention to this shift of interpretation is what allows a one-to-one mapping of premisses to diagrams. While each characteristic diagram can show all consistent types, diagrams representing the two different existential premisses can be distinguished by the types they establish. This representational scheme corresponds to the strategy of choosing the 'weakest' case, an extremely important principle in any proof system. In fact this is the same strategy as that adopted for choosing Berkeley's geometrical figures mentioned above. If exemplars can be ordered on a dimension of 'strength' defined so that what follows for any weaker case follows for all stronger cases, then the validity of an inference can be established by reasoning about weakest cases (one principle underlying some case-based reasoning). Abstraction is achieved by representativeness.

Finally, there is a degree of animation in our graphical system. Having superimposed the B circles we consider a range of relative positions for the A and C circles—our constraints are constraints on possible *motion*. We might represent this range by a set of dotted circles, exploiting a common convention for depicting motion in still diagrams. Animation here represents a disjunction of possibilities by a sequence of alternatives thought of as distributed

through time. The animation required for our purposes is less than the truly continuous variety. In each case, we only really need consider the five possible alternatives for the relative positions of the A and C circles.

Animation (either by real dynamic graphical representation, or by the suggestion of movement in a still representation) is another device for achieving abstraction. A primitive interpretation of a sequence of animated states has all the specificity of the primitive interpretation of a still graphic, but our proposed interpretation of the animation treats the sequence as a set of disjoined clauses and is thus another way of introducing abstraction. Animation is a particularly powerful technique when a substantial part of the scene stays constant. It is then perceptually easy to focus on the part that is changing and to exploit the redundancy of what remains the background. In the case of our diagrams, the motion can always be limited to one of the three circles.

This review of the conventions and strategies which introduce abstraction into the interpretation of graphical representations illustrates how our graphical system has a particularly transparent relation to the model theory of the syllogism. Validity is defined as truth of conclusion in all models consistent with premisses. Our diagrams represent all types consistent with two premisses and therefore represent all the available ingredient types out of which interpretations of the premisses which make them true must be constructed. The diagrams also distinguish types which must exist in all models (obligatory constituents of all models of the premisses) and therefore allow formulations of valid conclusions. This transparent relation between the diagrams and the definition of validity is a powerful aid to understanding the distinction between contingent truth and logical validity.

The discourse of proof in a deductive logic is a rather unusual type of discourse, and teachers of logic are only too familiar with the difficulties those new to logic experience in grasping the distinction. Indeed, even psychologists running experiments on logically naive subjects frequently experience the same phenomenon—'Do you want to know whether the conclusion *is* true or whether it *must* be true?' is a common question. The evidence, albeit based on the folk-lore of the profession rather than controlled experiment, is that the graphical methods outlined here make this relation transparent in a way that directly teaching a calculus obscures.

6 Implementing Euler's Circles in the Mind

In this section we consider issues raised by implementing our graphical algorithm for syllogism solution. It is important to be clear about the goal of implementation. Euler's Circles, embodied as rings, pastry-cutters and drawing pins, or pencil and paper drawings, together with the control mechanisms of the user, are a computational mechanism. We have shown how that mechanism can be interpreted as implementing syllogistic reasoning. Now we move on to questions about what computational mechanisms might implement this external mechanism in a syllogisers mind, in particular, when he or she solves syllogisms without recourse to external diagrams.

It iss worth noting that the same general questions can be asked about the syllogiser who uses pencil and paper to draw Euler's Circles: there is still a question about what computational

mechanisms operate in their minds. But it is a different question, and it differs chiefly in the lesser involvement of working memory for attribute bindings. Our reasons for pursuing the case where no external diagrams are used are firstly that we don't know very much about how people perform with externalised diagrams. The studies such as Newstead (1989) which have used drawings have mostly concentrated on subjects' interpretational errors rather than their ability to combine premisses. But secondly, it is the memory for bindings which is of particular interest. What distinguishes 'verbal reasoning problems', which people exhibit such difficulty with, from 'text comprehension', which people show such amazing facility at, is not the logical complexity of inferences involved. Understanding texts involves extremely complex inferences as witnessed by the attempts to simulate these inferences in AI. What distinguishes these areas is that so-called verbal reasoning problems require the establishing of temporary bindings of attributes to individuals without long term memory support, and the consideration of (possibly many) alternative patterns of binding of the same material. This is what presents such problems to the faculties of human deductive reasoning. We require an account of how the bindings that external Euler's Circles keep track of are implemented in the mind. And the account must explain the peculiar profile of abilities we observe.

The role that the graphical algorithm plays in this approach is that it presents different implementational problems than, say, mental models or natural deduction algorithms. These chiefly differ in the role that geometry plays in defining operations on circles. Since we observe that the external circles are conducive to reasoning, we speculate that it is because this external aids maps onto internal structures and processes perspicuously. And so the algorithm provides hints as to what these processes might be. We do not however believe, in the style of some imagery researchers, that what is implemented within is isomorphic to the full detail of the external aids. We rather look for minimal internal implementations, and these turn out to be much more modest than the registration diagrams on the printed page.

In order to be clear about the goals of implementation, we begin by reviewing some general phenomena of human verbal reasoning and the role PDP implementations may play in explaining them. We then discuss in more detail one particular PDP model of the mechanism of attribute binding in human working memory and the part this proposed mechanism might play in implementing the graphical algorithm. Finally we compare this model to other proposals relevant to explaining somewhat different aspects of human reasoning performance.

At the greatest level of generality, graphical representations and connectionist mechanisms share some salient computational properties. Graphical representations force the resolution of referential and spatial relations onto systems which may or may not possess the necessary information—they are 'specific' in our terminology. Connectionist mechanisms resolve multiple constraints which may logically only partially determine an outcome into fully determinate outputs. So, for example, a picture of two individuals cannot avoid determining to which individual a represented property is attributed. Similarly, a connectionist representation of this property binding will always resolve attributions one way or the other (see, e.g., the system described below).

Still remaining at this very general level of description, both graphical representations and connectionist mechanisms have the property that they *agglomerate* information—one representation is constructed which incorporates all fragments of information and therefore forces the resolution of relationships between constituent parts (see Stenning (1991) for a more ex-

tended comparison of agglomerative vs. analytical representations). Human verbal reasoning exhibits several phenomena which can be described as examples of this tendency to agglomerate information into total rather than partial representations. Both the literature on n-term series problems (e.g. Huttenlocher 1968) and that on syllogisms (e.g. Johnson-Laird 1983, Erickson 1974) make this observation, as does the text memory literature (e.g. Kintsch & van Dijk (1978) in which total connectivity is often taken as definitional of coherent text.

However much these general comparisons may lack in implementational detail, their importance cannot be overestimated. Conventional rule based approaches to human reasoning may have provided great implementational detail but they have never offered any *explanation* of this pervasive property of human reasoning strategy. Nor is clear that they could. The generalised computational power of conventional rule-based approaches is their downfall in this respect. It is as easy for them to compute with partial as with total representations.

Another important general property of human deductive reasoning is that it is strongly affected by the content over which people reason. Content effects take place at many levels. Perhaps the most straightforward is interference by premisses or conclusions which are known to have a particular truth value in the real world, and particularly if this truth value has heavy emotive undertones (e.g. Janis & Frick 1943, Lefford 1946, Oakhill, Garnham & Johnson-Laird 1990). Another type of content effect is that in which content affects the interpretation placed on the premisses of an argument. This type of effect is best illustrated in the conditional reasoning literature (e.g. Wason 1968). Yet another type of content effect, the one that will chiefly concern us here, is an effect of richly associative content *enhancing* performance as compared to performance with schematic symbolic material. As Johnson-Laird (1983) points out, artists, beekeepers and chemists are easier to reason about syllogistically than As, Bs and Cs.

It seems unlikely, at least to the present author, that the first two types of content effects are particularly suited to an explanation in terms of an underlying PDP architecture as opposed to a conventional one. At the very least, they first require a theory about how content leads to different processing choices which could perfectly well be expressed in conventional rule-based terms. The last type of effect does however seem to be a good candidate for distinguishing underlying architectures. The phenomenon consists of the observation that material with rich associative history is mnemonically more manageable in whatever working memory it is that supports syllogistic reasoning. Associative memory is just what can be modelled more adequately in a PDP framework than a conventional one.

However, the modelling problem is an interesting one, because the memory task involved in solving syllogisms 'in the head' is not merely one of associative memory. In common with most deductive tasks, it involves creating representations of hypothetical individuals from combinations of properties, and considering various different assortments of these individuals. This binding and rebinding is just what is not easy to implement in connectionist systems although there is a growing list of proposals about how this might be achieved (e.g. Barnden 1989, Halford et al. (this volume), Shastri & Ajjanagadde 1989, Stenning & Levy 1988). So modelling the impact of associative content on verbal reasoning is a nicely balanced exercise for PDP modellers.

These alternative proposals for solving the binding problem are each designed for rather different purposes and from rather different perspectives. Stenning and Levy's proposals are

motivated by the desire to fit particular data that arise in a memory-for-bindings task whereas the other proposals are all for extensions to the PDP modeller's armoury of architectures. It is not clear whether or not these other proposals could exhibit the sort of error-under-noise behaviour which the Stenning and Levy model simulates because they have not addressed this issue.

Stenning and Levy took the experimental data of Stenning, Shepherd & Levy (1988) consisting of frequencies of different types of errors of recall in a task where subjects remembered mappings of properties onto individuals. They showed that this data could be modelled by assuming a distributed representation of binding, and indeed required such an assumption. We will briefly describe the model they proposed. Our purpose is to explore the implications of their findings for the working memory underlying syllogisms. Euler's Circles, because they are graphical, force the specification of every individual on all the represented property dimensions—if there are three circles on a plane, every region of the plane is either inside or outside each circle. Stenning and Levy's model shares this computational property, not because it is graphical, but because it represents bindings through contraint satisfaction.

Stenning, Shepherd and & Levy's data come from a task in which subjects read paragraphs attributing four properties each to two individuals (objects with shapes, colours, textures and sizes, or people with professions, nationalities, physiques and temperaments). The data which concern us here consist of frequencies of twenty categories of error selected for what they reveal about the underlying representation. The initial observation was that some multiple attribute errors (cases in which more than one property is incorrect) are much more common than they should be if each attribute of each individual were independently bound. For example, a common error is to remember that the two individuals contrasted on a dimension, but get the binding the wrong way round—a Polish bishop and a Swiss dentist are retrieved as a Polish dentist and a Swiss Bishop. This is much commoner than remembering two Polish individuals as two Swiss ones, and is about as common as simply making one error on a dimension on which the individuals contrast—a Polish bishop and a Swiss dentist retrieved as a Polish bishop and a Polish dentist.

Such observed frequencies suggest that there is something in common between the representations of the two 'reversals' which makes them more similar and therefore more confusable than independent non-redundant representations would be. These isolated observations pose a general question—is there a set of 'features' which can be represented independently (and therefore corrupted in memory independently) which can capture the frequencies of the errors which define combinations of their corruption? Features simply state facts about the pair of individuals (e.g., 'There are two people of same nationality', 'There is a person both happy and Polish' etc.) and so their truth value is determined for any pair of people presented. Any error (best thought of as a transformation from input description to recall description) therefore changes the truth value of some features and leaves others unchanged. The more features which change, the less likely is the error. The features may be weighted to reflect their salience. This methodology can be seen as another application of principles underlying such classic perceptual feature extractions as Miller & Nicely's (1955) demonstration of phonemic features underlying consonant confusions in speech.

Stenning, Shepherd & Levy show that a set of 15 features which fit the error frequency data well can be found using regression techniques. Table 3 shows an example feature set with

Feature	Coefficient	Standard Error
Intercept	-4.72	
$\exists$ x($\neg$Bx&$\neg$Dx)	0.23	0.11
$\exists$ x($\neg$Ax&$\neg$Dx)	0.26	0.11
$\exists$ x($\neg$Ax&$\neg$Cx)	0.38	0.10
$\exists$ x($\neg$Bx&Cx&Dx)	0.43	0.12
$\exists$ x(Ax&Bx)	0.47	0.12
$\exists$ x($\neg$Bx&$\neg$Cx&$\neg$Dx)	0.58	0.12
$\exists$ x(Ax&Bx&$\neg$Dx)	0.25	0.10
$\exists$ x(Ax&Cx&$\neg$Dx)	0.34	0.10
$\exists$ x($\neg$Ax&Bx&$\neg$Cx)	0.43	0.11
$\exists$ x($\neg$Ax&Bx&$\neg$Cx& Dx)	0.50	0.12
$\exists$ x(Cx)	0.68	0.09
$\exists$ x$\exists$ y(Ax&$\neg$Ay&x$\neq$y)	0.90	0.15
$\exists$ x$\exists$ y(Bx&$\neg$By&x$\neq$y)	0.21	0.09
$\exists$ x$\exists$ y(Dx&$\neg$Dy&x$\neq$y)	0.73	0.07
nmat	0.22	0.07

Table 3: Summary of features in representation of binding (adapted from Stenning, Shepherd & Levy 1988). The coefficients are measures of the weight attached to features in predicting error frequencies. 'Nmat' is an abbreviation for 'There is more than one mismatched dimension'.

their relative weightings. This set illustrates the web of logical relations between features. Once corruption occurs (the truth value of one or more features changes) the likelihood is that the set will then be inconsistent (there are only 136 consistent sets of values out of 2^{15}). The memory retrieval mechanism then has the constraint satisfaction problem of finding a closest fitting pair of individuals which maximises the number (or aggregated salience) of feature values in the corrupt representation. This is a hard problem of just the type that PDP networks perform well. Stenning and Levy show that a simple backpropagation network can learn the logic relating the features in the regression model from presentations of well-formed inputs with their corresponding outputs. When the trained network is then presented with randomly corrupted inputs (all of which are novel stimuli), its output exhibits a frequency profile of error types which closely approximates both the regression model and the human data. Figure 6 shows the network and Figure 7 compares the frequency profiles of human memory data, the regression model's predictions, the PDP simulation, and a non-redundant representation of the same information. Performance of this non-redundant representation contrasts with the distributed memory simulation (and with human memory) most clearly in producing no triple- or quadruple-property errors. These complex errors are relatively frequent in distributed memories.

What this model achieves is a reduction of the complex binding problem posed by the experimental task into a simpler one of binding attributes within the simple conjunctive existential features in the regression model. (We would claim that the feature 'Both A or both B' need not be expressed disjunctively). Knowledge of co-instantiations of properties establishes knowledge of references. A more complete model would have to explain how the simple binding problem within the existential features of this model is implemented in terms of associative

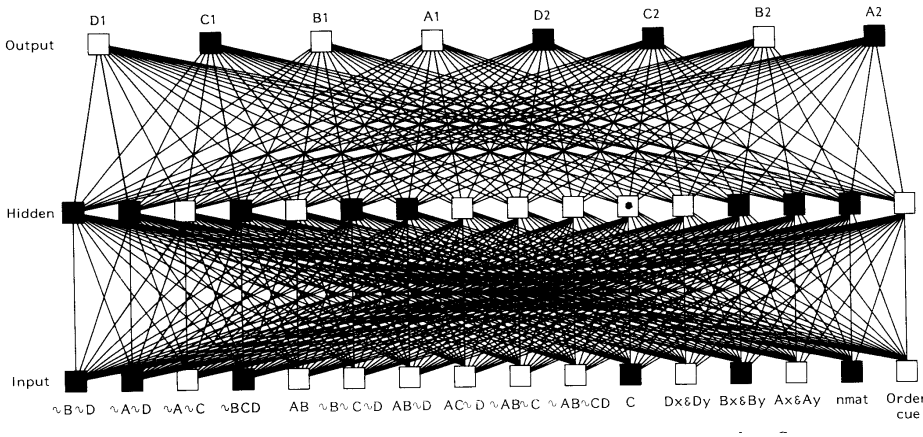

Figure 6: A PDP network which retrieves the bindings of four binary properties to two individuals. The input layer (bottom) is a distributed representation of the set of bindings in the output layer (top). From Stenning and Levy 1988, by permission of the publishers, Butterworth-Heinemann Ltd ©.

memory for content. We do not have a simulated model of this implementation but it is worth making some comments about the nature of the solution required because this has implications for the explanation of the content dependence of memory.

We assume that a feature such as 'There is someone both Polish and a bishop' may be further analysed in memory into some contentful associations which enable the subject to choose between the possible bindings of these properties, namely: Polish bishop, Swiss bishop, Polish dentist and Swiss dentist. So, for example, if a subject remembered there being evidence of a catholic person, that might be sufficient to implement the binding between 'polish' and 'bishop'. Swiss bishops aren't stereotypically catholic; polish dentists aren't stereotypically religious; and Swiss dentists aren't stereotypically either. (If this example does not work for you, choose whatever example associations do work for your knowledge/beliefs about these properties). Similarly, the evidence is that the implementation of the feature that there are two people of different nationality is achieved in contentful fashion. Memory for matching/mismatching on dimensions does not deteriorate with proactive interference in the way one would expect if the coding were simply a schematic 'match' or 'mismatch'.

The computational model is therefore incomplete but nevertheless performs an important reduction of the problem from the task of binding attributes to their references to the problem of conjoining properties in simple existential statements. The combination of experimental observations and modelling provide strong evidence that binding is represented in a distributed contentful fashion and resolved by constraint satisfaction inference from such a representation. We now explore the implications of this architecture for verbal reasoning and for graphical representations, and compare this proposal with other solutions to the binding problem.

As we have seen, solving syllogisms by the graphical algorithm we have specified requires keeping track of collections of types of individual defined in terms of the three properties

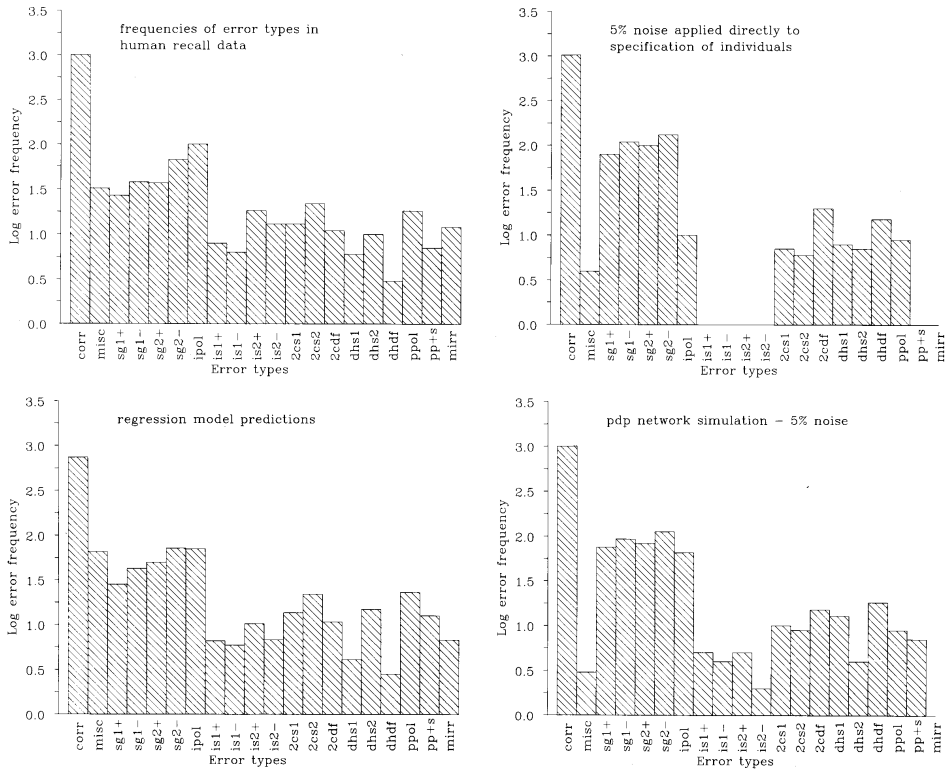

Figure 7: Log frequencies of twenty categories of error for (a) human data, (b) 5% noise applied directly to descriptions of individuals, (c) regression model predictions, and (d) PDP network simulation—5% noise. Details of error-types are explained in Stenning, Shepherd & Levy 1988. From Stenning and Levy 1988, by permission of the publishers, Butterworth--Heinemann Ltd ©.

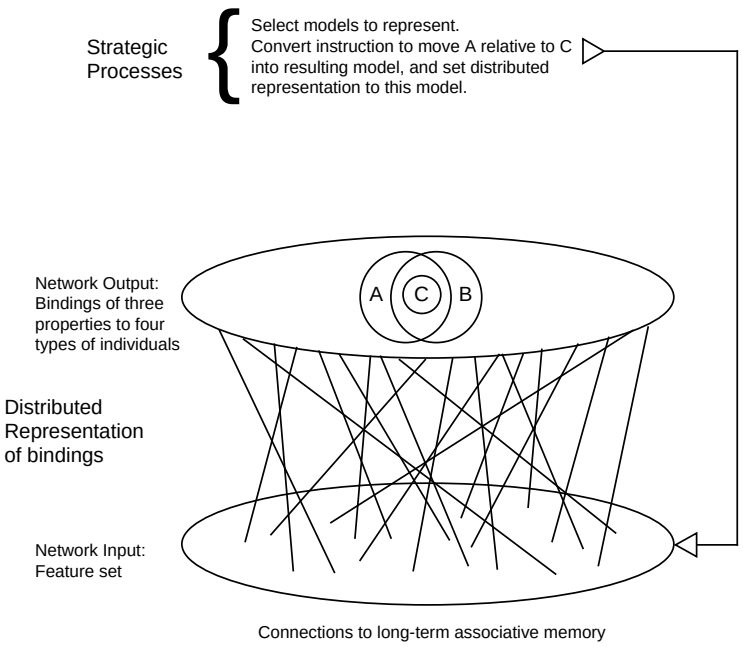

Figure 8: Schematic embedding of a distributed memory for bindings in a syllogism solution system.

contained in the premisses. This core memory problem is closely analogous to the binding problem set to Stenning, Shepherd and Levy's subjects. It is not the only memory required to implement syllogistic reasoning—subjects also need to keep track of some of the information about premisses, and require internal memory of the strategies and processes of reasoning. Nevertheless, this memory for bindings is a central part of the difficulty of difficult syllogisms for human reasoners. So the line of investigation we will pursue here is to suppose that the heart of the syllogism solving system is a memory for bindings, based on a mechanism like Stenning and Levy's model. The strategic processes of reasoning then operate over the output of this module (the collection of types represented on the output units) and have the ability to reset the model held by the module by changing its inputs. Figure 8 shows a schematic architecture. What syllogistic reasoning shares with the straight memory-for-bindings task is that models consist of collections of individuals (in this case defined by three properties rather than four). What syllogistic reasoning adds to the straight memory-for-bindings task of Stenning, Shepherd & Levy is the need to consider several patterns of binding in one problem, at least in the difficult cases. To see what is involved, consider the registration diagrams in Figures 3, 4 and 5. Each diagram represents collections of types directly, and also implicitly represents ranges of collections which correspond to potential movements of the A and C circles. Consider first what is involved in representing the basic static registration diagrams.

Information about the number of individuals present plays an important role in the Stenning and Levy model of binding in terms of existential features. Only if we know that there are

exactly two individuals can we make the necessary inferences from the state of the feature set. Feature sets could be constructed for different numbers of individuals, but having a determinate number is important. Solving the existential constraints into bindings relies on knowing how many indiviuals there are in the model. If human working memory implemented the external graphical representations completely, that would require representing collections of from three to seven types of individual. Interestingly, this range is smaller among syllogisms which do have valid conclusions (from three to five types). The only ones having six or seven types in their registration diagrams are ones with two particular premisses, and these are syllogisms for which subjects readily reject any conclusion as valid. So one might suppose that an application of this principle applied to show that no diagram-analogue need be constructed. This would leave the need for the representation of three different sized sets of types: three-, four- and five-membered.

Another strategy that can be used to reduce the complexity of the binding problem is to focus on the nailed sub-area of the diagram. This focus reduces the number of types to a constant singleton set but would still require some representation of constraints on the movement of A and C circles. This brings us to consideration of movement. This in turn suggests a further strategy for limiting the variation in number of individuals represented. Movement is the second extension that would be required to implement our graphical algorithm in a binding mechanism like that of Stenning and Levy.

Each registration diagram is an invitation to consider whether or not the A and C circles can move relative to each other. Nails are constraints on such movement. Externally we could actually implement Euler's Circles as rings (preferably expandable and contractable rings) sliding on a flat surface constrained by nails. But how is discovery of constraint on movement to be implemented internally?

Movement consists in change of model i.e. change of membership of the set of types. But not any change from model to model is possible—the possible changes are structured by the geometry of the circles. We can conceptualise these movements as transitions between models arranged in a seven dimensional space, where each dimension represents the presence or absence of a type in the model. One of the main functions performed by the graphical representations is that they make it transparent which transitions are possible—by merely considering a direction of relative movement of the A and C circles, it becomes clear which type(s) of individual will be added to the model or deleted from the model by the movement.

The graphical aid reveals an important property of the transitions—that they always add or delete just one type. The transitions are minimal changes. Or more precisely, any transition that makes more than one addition/deletion of types can be replaced by several transitions which change only one type. The graphical representations introduce a certain continuity into the range of models to be considered.

It is this continuity which affords important mnemonic savings. The positive or negative nature of the syllogism determines the relevant direction of movement which the reasoner has to consider. Movement in that direction can be a minimal movement which will add or delete just one type. The distribution of shading will allow the choice of which is the relevant transition to consider if there is more than one. So this apparatus offers the possibility of considering just one model (represented by the registration diagram of the syllogism), and just one addition or deletion of a type to or from that model. This reduces the binding problem

to the holding of one model—one collection of types. The identity of the addition/deletion to be considered can be held in some other memory, or as an annotation on the main memory for the registration diagram. Instead of having to consider perhaps three whole models, each arbitrarily related to the others, one model with potential minimal change suffices.

How might this movement mechanism be implemented? The minimal supposition would be that some rule based system encapsulating the possible transitions between models could operate as an external strategic controller of a binding mechanism of the type considered above. Such a controller could reset the model held in the PDP mechanism by recourse to changing the state of its inputs. In such an architecture, the geometry of the graphics is built into a module quite separate from the binding mechanism which merely holds arbitrary collections of types. The content dependence of the binding mechanism would still have its effects, but there would be no special influence of content on the transitions from model to model.

A more radical direction for investigation would be a connectionist implementation of the possible transitions from model to model which correspond to circle movement. Such a system would be a connectionist implementation of the rule based controller. It would have to be capable of representing each of the models corresponding to a registration diagram (including all those created by movements of the A and C circles), and representing directions of movement (at least 'A towards C' and 'A away from C'). Starting in a state representing one model, and given an input corresponding to a direction of movement, the system would have to transition into the state representing the next model in that direction. A sequence of such direction commands issued from outside the system would cause it to travel through the space of registration diagrams.

This would require an implementation in a dynamic connectionist system analogous to, but more complex than, that of Amit (1987) or Dehaene, Changeux and Nadal (1987). One might expect systems like this (and with much greater complexity) to exist for the purpose of planning action on visually presented scenes. Although little is known about how such systems can be configured by learning algorithms, it is not clear that biological systems do arise in this way. It seems likely that much preconfiguration of these visual/motor mechanisms has been achieved through evolution.

Several literatures are relevant to our abilities to manipulate spatial configurations without external aid. There are studies of tasks designed to demonstrate the *existence* of specialised mechanisms for such manipulations (e.g. Brooks 1968) and there is extensive use of related items in psychometric *measurement* of 'spatial' ability. There is also the literature on the visuo-spatial sketchpad (Baddeley 1986) responsible for visual working memory. and the literature on memory for information about static spatial arrays described in text (e.g. Mani & Johnson-Laird 1982). The 'mental rotation' task of Shepherd & Metzler (1971) may involve some of the relevant mechanisms. But none of the tasks involved explicitly analyse our ability to make judgements of relations between components of spatial arrays that would result from actions on initial configurations. In this connection, Hinton (1979, 1980) makes the interesting argument that it is the type of continuity of representation observed in the animations of our registration diagrams that plays a major part in explaining data on visualisation, but that this sort of continuity does not presuppose what he calls 'atomic depiction' which is often claimed to be required by observed mental rotation data. The contrast is roughly between

structural descriptions which may have continuously valued variables attached to them, as opposed to genuinely analogue representations.

A dynamic connectionist controller which implemented the operations on registration diagrams could be separate from the memory for bindings module (in the same way that the rule based controller might). The 'names' used to identify models inside the controller module would not necessarily have to be analysable down to the level of the component properties of their constituent types, and the problem of binding together properties would not arise within the controller. It would simply operate on the output of the memory-for-bindings module.

The remaining issue is whether memory-for-bindings and controller should be further integrated into one module. The answer will make a difference to the type of error behaviour the system will exhibit. This is really a question of how far the content-dependent memory for bindings ramifies through the system. If the separate controller simply operates on the output of the binding module, one would expect content to affect general memorability of the base registration diagram from which reasoning proceeds, but not to affect which models are considered as deformations of this model, corresponding to movements of the A and C circles. If the memory used by the controller is the same memory as the memory-for-bindings, and is therefore content dependent in the same way, one would expect content to affect the ease of considering movements of the A and C circles.

To make this distinction more concrete, it can be posed with regard to the user of external graphical aids. If a reasoner translates a syllogism into the medium of three circles simply tagged by letters, reasoning proceeds by considering the circles without recourse to the content of the premisses save as it is represented by the schematic letters, and then the results are reinterpreted into contentful terms, this is analogous to the controller being separated from the memory for bindings. If, however, the encoding into the graphical representation maintains content which has effects through its associative properties on the manipulations performed on the circles, then the two modules are integrated, and the interference effects will be different. Without detailed simulations, and more detailed error data than is currently available these alternatives cannot be decided. A further complexity is that reasoners of different levels of expertise exhibit different sorts of content effects and may best be modelled by different systems.

Finally, it is useful to compare this line of development of implementations for our graphical algorithm with other proposals for attribute binding within connectionist systems. CONPOSIT, Barnden's (1989) system of binding exploiting Relative Position Encoding (RPE) and Pattern Similarity Association (PSA) actually chooses Johnson-Laird's mental model method of syllogistic reasoning as an illustrative domain. As mentioned above, the purpose of this implementation is to illustrate connectionist techniques which are of considerable cognitive interest in themselves, but the issue of the content dependence of human reasoning is not explored, largely perhaps because it is not explored in the mental model theory chosen for implementation. RPE solves the problem of setting up temporary and arbitrary bindings of attributes together by placing tokens of the attributes in adjacent positions in a spatial array representation. What is adjacent is bound. What is distant is not bound. From one perspective this constitutes 'symbol processing', but from another it can be seen as implementing temporary virtual nets.

In the current context of spatial analogy, this proposal is intriguing because it uses a spatial

representation in which adjacency in a 2-D sheet of elements carries the binding information. It appears to use the sheet in a quite different way than Euler's Circles, though we have already mentioned above that mental models notation hides its relation to Euler's Circles. Like mental models, CONPOSIT uses RPE to represent types of individual as clumps of tokens in adjacent areas, but does not exploit the properties of the plane to relate different collections of types through movement in the plane. Whether or not further exploitation of the spatial arrangement is possible is at present an open question, as is the question of how content effects can be modelled in this framework.

Other current connectionist proposals for implementing binding have not been so directly related to syllogistic reasoning by their authors. Shastri & Ajjanagadde's (1989) implementation of binding by phase-locking is designed to implement a data-base query language for performing 'reflex' reasoning—that fragment of reasoning over the contents of a data-base which can be performed quickly, and with some degree of independence of the size. This system implements queries of the form of an impressively large fragment of the predicate calculus, and does so in time linear with the shortest proof. This system has several properties relevant to human reasoning performance. It has severe limitations on how many individuals it can reason about (variables bound during an episode of reasoning) but can reason over very large data-bases of long term information. This accords well with human performance—in fact the observation is so pervasive that it is scarcely mentioned in the literature.

Although the phase locking implementation of bindings may be applicable to our domain, there are a number of obstacles in the path of applying Shastri and Ajjanagadde's current system to syllogistic reasoning. The distinction between long-term database and an episode of reasoning over it does not easily transfer to the syllogistic case in which the domain of reasoning and general facts about it are created anew for each problem. The very facility of human reasoning over the sort of problem Shastri and Ajjanagadde's studies contrasted with the laborious and unreliable performance with some syllogisms suggests that at least some different systems are involved. In fact, the system which performs inferences over the long term knowledge base is one of the systems which actually leads to interference of content with reasoning in the syllogistic task. If people could isolate their knowledge that 'All men are women' is false from their syllogising system, they would not display the interference effects psychologists observe in reasoning with such premisses. Although the systems are not identical, they also do not seem to easily be totally decoupled.

Shastri and Ajjanagadde's proposals, like those of Barnden, do not address the issue of interference effects: that is not the focus of Shastri and Ajjanagadde's work and the systems do not use the distributed representations which would bring this issue to the fore.

To summarise this discussion of connectionist implementation of our graphical algorithm, two broadly contrasting interests in the problem have been the desire to extend the computational abilities of connectionist systems on the one hand, and to model features of human performance, particularly the intrusion of content into reasoning, on the other. The two aims have produced proposals which are largely complementary rather than competitive, and considerable work remains to be done to see how the two sets of results can be integrated. What is important to Stenning and Levy's concerns is that the binding of attributes to individuals that appear in problem descriptions is based on a constraint satisfaction inference from data at a lower level of representation. It is this distinction of levels which opens the

door to studies of the content dependence of working memory and its impact on reasoning. The particularities of the constraint-satisfaction device that appears in their model is of less concern. It is possible that the other proposals for binding described here may be combined with distributed representations that can be connected to a long term-associative memory, and if this proves to be the case the two broad aims will be met in one system.

7 Analogy and the reduction of expressiveness

Our case study has shown how one detailed working out of the analogy of spatial containment in the form of Euler's Circles is less expressive than elementary set-theory. The spatial analogy cuts down the problem space. Perhaps the most instructive conclusion that can be drawn from developing the full detail of this example for which we know the destination task is that it is the combination of vehicle and topic that results in this reduction in expressive power. It is not that spatial containment necessarily less rich than set membership, but it is bringing the general analogy to bear on the use of circles to model syllogisms that achieves the gains in processability. The general analogy's applicability is reliant on the limitations of the logic which it is used to model as much as on the geometrical limitations of circles, and on the combination of logic and geometry in particular.

Even if we augment syllogistic logic in quite trivial ways we have to augment our graphical resources beyond circles. If we move to full monadic predicate calculus, even if we restrict our vocabulary to three predicates, we can no longer manage with closed curves. The relevant addition here is of the machinery of iterated quantification and a full set of connectives. This addition makes it possible to distinguish every model from every other (as mentioned above) and therefore to require diagrams for all the models without Euler's Circle diagrams (see Figure 3). Some of these models require discontinuous regions. Some of the complexities that arise with increasing the number of predicates beyond three are illustrated in Lewis Carol's text through his correspondence with Venn (see Dodgson (1896)).

Once we go beyond monadic predicate calculus to polyadic calculi, graphical methods have to retrench still further in what they can achieve. There are other ways of using graphical methods to teach polyadic predicate logic (cf. Barwise and Etchemendy's Tarski's World and Hyperproof programs), but they do not exploit the spatial inclusion analogy. Even for monadic predicate calculus, the analogy is unhelpful because it is too inexpressive.

Does analogy generally work by limiting expressive power? Is a muzzle a good analogy for analogy? Our temptation is to answer yes. At least this seems a hypothesis worthy of pursuit. In the cases of analogy most discussed in the literature it is often hard to see exactly what the end result inference system is that corresponds to the Euler's Circle system for syllogism solution. But it does seem that these end results may exist implicitly. When plumbing is used to understand electricity, it is only a small part of electrical behaviour which the analogy models (e.g., Ohm's Law, but not magnetic field induction). Along with this conclusion that what is derived from a general analogy is often a circumscribed inference system comes some clarification of the relation between analogy and general issues of representation from which both fields may benefit.

What does connectionist implementation add to these morals about analogy? Connectio-

nism compared to symbolic approaches is analogous to graphical representations compared to linguistic (logical) representations—the key to understanding their contribution to cognitive science is their curtailment of expression. In our specific proposals about connectionist implementations of our graphical algorithm the implementations play two roles: the first is to analyse attribute binding as an inferential achievement (rather than a representational primitive) in a way that explains why people choose a reasoning strategy which involves whole patterns of binding. This in turns holds out the future possibility of explaining the observed differences of performance when people reason with As, Bs and Cs and with artists, beekeepers and chemists because it suggests how to connect content effects with binding performance.

The second role of connectionist implementation is that it can exploit the continuity afforded by graphical representations in simulating the movement of circles. If we have apparatus 'designed' for imagining spatial rearrangements of a certain continuous variety, then this could be brought to bear on graphical representations in a way that it could not be brought to bear on other notationally equivalent representations (e.g., mental models). In doing so it could effect mnemonic savings which would bring the syllogism within the compass of a memory-for-bindings for a fixed number of individuals.

References:

Amit, D. J. [1987] Neural Networks Counting Chimes, Technical Report RI/87/49, Racah Institute of Physics, Hebrew University, Jerusalem

Baddeley, A. [1986] *Working Memory.* Oxford: Oxford University Press.

Barnden, J. [1989] Neural-net implementation of complex symbol processing in a mental model approach. In *International Joint Conference on Artificial Intelligence-89*, 1989, pp568-573.

Barwise, J. and Etchemendy, J. [1990] Information, Infons and Inference. Chapter 2 in Cooper, R. (ed.) *Situation Theory and its Applications*, Volume 1.

Brooks, L. [1968] Spatial and verbal components of the act of recall. *Canadian Journal of Psychology*, **22**, 349- 368.

Berkeley, G. [1937] *The Principles of Human Knowledge*: *the text of the first edition (1710).* London: A. Brown.

Dehaene, S., Changeux, J. and Nadal, J. [1987] Neural networks that learn temporal sequences by selection. *Proc. Natl. Acad. Sci. USA*, **84**, 2727-2731.

Dodgson, C. [1896] Symbolic Logic. Book III in *Lewis Carroll's Symbolic Logic*: *edited by W. W. Bartley.* Hassocks, Sussex: Harvester Press.

Erickson, J. R. [1974] A set analysis theory of behaviour in formal syllogistic reasoning tasks. In Solso, R. (ed.) *Loyola Symposium on Cognition*, Volume 2. Hillsdale, N.J.: Lawrence Erlbaum Associates.

Funt, B. V. [1977] WHISPER: a problem solving system utilizing diagrams and a parallel processing retina. In *International Joint Conference on Artificial Intelligence*, 1977, pp459-64.

Funt, B. V. [1980] Problem-Solving with diagrammatic representations. *AI*, **13**, 210-230.

Gelernter, H. [1963] Realization of a geometry-theorem proving machine. In Feigenbaum, E. A. and Feldman, J. (eds.) *Computers and Thought*, pp134-152. N. Y.: McGraw-Hill.

Hinton., G. [1979] Some Demonstrations of the Effects of Structural Descriptions in Mental Imagery. *Cognitive Science*, **3**, 231-250.

Hinton, G. [1980] Frames of reference and mental imagery. Chapter 15 in Long, J. and Baddeley, A. (eds.) *Attention and Performance*, Volume IX, pp261-277. Hillsdale: Erlbaum.

Holland, J., Holyoak, K., Nisbett, R. and Thagard, P. [1986] *Induction.* Cambridge, Mass.: MIT Press.

Hunter, G. [1971] *Metalogic*: *an introduction to the metalogic of standard first-order logic.* London: Macmillan.

Huttenlocher, J. [68] Constructing spatial images : a strategy in reasoning. *Psych Rev*, **75**, 550-60.

Janis, I. L. and Frick, F. [1943] The relationship between attitudes towards conclusions and errors in judging logical validity of syllogisms. *J. of Experimental Psychology*, **33**, 73-77.

Kintsch, W. and Dijk, T. A. [1978] Towards a model of text comprehension and reproduction. *Psychological Review*, **85**, 363-394.

Lefford, A. [1946] The influence of emotional subject matter on logical reasoning. *J. of General Psychology*, **34**, 127-151.

Levesque, H. [1988] Logic and the Complexity of Reasoning. *Journal of Philosophical Logic*, **17**, 355-389.

Lindsay, R. K. [1988] Images and inference. *Cognition*, **29**, 229-250.

Lyons, J. [1968] *Introduction to Theoretical Linguistics.* Cambridge: Cambridge University Press.

Mani, K. and Johnson-Laird, P. N. [1982] The mental representation of spatial descriptions. *Memory and Cognition*, **10**, 81-87.

Newstead, S. E. [1989] Interpretational Errors in Syllogistic Reasoning. *J. of Memory and Language*, **28**, 78-91.

Miller, G. A. and Nicely, P. [1955] An analysis of perceptual confusions among some English consonants. *Journal of the Acoustical Society of America*, **27**, 338-352.

Oakhill, J., Garnham, A. and Johnson-Laird, P. N. [1990] Belief bias effects in syllogistic reasoning. In Gilhooly, K. (ed.) *Lines of Thinking*, Volume 1. London: Wiley.

Piaget, J. and Inhelder, B. [1956] *The child's conception of space.* London: Routledge and Kegan Paul.

Richards, I. A. [1936] *The Philosophy of Rhetoric.* London: Oxford University Press.

Shastri, L. and Ajjanagadde, V. [1989] A Connectionist System for Rule Based Reasoning with Multi-place Predicates and Variables. Technical Report No. MS-CIS-89-06, Department of Computer and Information Science, School of Engineering and Applied Science, Philadelphia, 1989.

Shepard, R. N. and Metzler, J. [1971] Mental rotation of three-dimensional objects. *Science*, **171**, 701-703.

Stenning, K. [1991] Distinguishing conceptual and empirical issues about mental models. In Rogers, Y., Rutherford, A. and Bibby, P. (eds.) *Models in the Mind.* Academic Press.

Stenning, K. and Levy, J. [1988] Knowledge-rich solutions to the 'binding problem': some human computational mechanisms. *Knowledge Based Systems*, **1**, 143-152.

Stenning, K. and Oberlander, J. [1991a] Reasoning with words, pictures and calculi: computation versus justification. In Barwise, J., Gawron, J. M., Plotkin, G. and Tutiya, S. (eds.) *Situation Theory and its Applications*, Volume 2. Chicago: Chicago U P.

Stenning, K. and Oberlander, J. [1991b] A cognitive theory of graphical and linguistic reasoning: logic and implementation. Research Paper No. 20, HCRC, Edinburgh University, Edinburgh, 91.

Stenning, K., Shepherd, M. and Levy, J. [1988] On the construction of representations for individuals from descriptions in text. *Language and Cognitive Processes*, **2**, 129- 164.

Wason, P. C. [1968] Reasoning about a rule. *Quarterly Journal of Experimental Psychology*, **20**, 273-281.

Yule, P. [1991] Experimental investigation of a novel syllogism task. MSc Thesis, University of Edinburgh.